ATLASELE LINGVISTICE CA RESURSE LEXICOGRAFICE: DE LA VERSIUNEA TIPĂRITĂ LA FORMATUL DIGITAL[1]

SILVIU-IOAN BEJINARIU[*], FLORIN-TEODOR OLARIU[**], RAMONA LUCA[*],
VERONICA OLARIU[**], LUMINIȚA BOTOȘINEANU[**],
TUDOR BUMBU[***], LUDMILA MALAHOV[***]

1. ATLASELE LINGVISTICE ÎN FORMAT TIPĂRIT – IZVOARE DE INFORMAȚII LEXICOGRAFICE

1.1. Importanța resurselor lingvistice de factură dialectală (atlase, texte orale în registru dialectal [= „texte dialectale”], glosare, dicționare etc.) pentru lucrările lexicografice este un fapt bine cunoscut, fiind adus în mod constant în prim-plan de numeroși lingviști de-a lungul timpului. Astfel, unul dintre primii specialiști/cercetători care au susținut în mod explicit importanța atlaselor lingvistice în procesul de redactare a dicționarelor a fost Sextil Pușcariu, într-o recenzie ocazionată de apariția primului volum din *Atlasul lingvistic român* (1938), unde, în calitate de coordonator atât al lucrării recenzate, cât și al *Dicționarului limbii române* – deci de foarte bun cunoscător al celor două opere fundamentale ale lingvisticii românești –, afirma: „În măsură mai mare decât pentru fonetică, fonologie, morfologie și derivațiune, materialul cuprins în Atlas e important pentru **lexic** [subl.a.]” (Pușcariu 1938: 425). În plus, în același loc, Pușcariu scotea în evidență caracterul convergent al celor două instrumente în abordarea faptelor de limbă:

> „Compararea Atlasului cu Dicționarul ne ajută adesea să găsim datele complementare, care lipsesc și într-unul și într-altul consultat singur. De aceea chestiunea «Atlas sau Dicționar?», cum a pus-o odată L. Spitzer, nu poate avea decât răspunsul: Atlas și Dicționar” (*ibidem*: 441–442)[2].

[1] Această lucrare a apărut cu sprijinul financiar al Ministerului Cercetării, Inovării şi Digitalizării, CNCS-UEFISCDI, în cadrul proiectului cu numărul PN-IV-P8-8.3-ROMD-2023-0318.

[*] Institutul de Informatică Teoretică, Academia Română – Filiala Iași, Str. T. Codrescu, nr. 2.

[**] Institutul de Filologie Română „Alexandru Philippide”, Academia Română – Filiala Iași, Str. T. Codrescu, nr. 2.

[***] Institutul de Matematică și Informatică „Vladimir Andrunachievici”, Universitatea de Stat din Moldova, Chișinău, Str. Academiei, nr. 5.

[2] Această opinie, un adevărat crez al lui Pușcariu, a fost enunțată deja cu câțiva ani mai înainte aproape în aceiași termeni: „În general, întrebarea «atlas sau dicționar» mi se pare oțioasă, căci răspunsul nu poate fi decât «atlas *și* dicționar». Cele două opere se completează una pe alta și dau împreună imaginea justă a limbei” (Pușcariu 1931: 505).

1.2. Ideea a fost adusă recent în discuție inclusiv de dialectologii din centrul academic ieșean, printr-o serie de articole care, elaborate în special din perspectiva procesului de reluare a redactării seriei DA a dicționarului-tezaur al limbii române, dar și a actualizării cu noi date dialectale a volumelor din seria DLR a aceleiași opere de referință a culturii române, abordează diverse aspecte ale relației atlas – dicționar. Problemele tratate în aceste articole vizează fie unele aspecte de natură „tehnică", precum necesitatea stabilirii unor reguli clare referitoare la preluarea, adaptarea și literarizarea materialului dialectal din proiectele ALR și ALRR/NALR – atât atlase, cât și volume de etno- și sociotexte – în dicționarul-tezaur (Dumistrăcel et al. 2015), fie scot în evidență statutul atlaselor regionale din seria ALRR/NALR (Dumistrăcel et al. 2011) sau al dicționarelor dialectale recente, cum este cazul *Dicționarului graiurilor dacoromâne sudice* (Hreapcă 2012), ca resurse lexicografice de primă importanță pentru DLR, mai ales din perspectiva bogăției de date pe care acestea le aduc – de multe ori în calitate de prime atestări – în atenția redactorilor dicționarului-tezaur[3].

Dat fiind acest fapt incontestabil, autorii în cauză nu numai că se poziționează împotriva ideii de a diminua importanța informației dialectale în redactarea articolelor din DLR – idee care a fost vehiculată în ultima perioadă în cercurile academice de specialitate –, ci vin și cu soluții în direcția valorificării optime a acestei informații, inclusiv printr-o regândire a sistemului de redactare a cuvintelor-titlu care să ușureze accesul la materialul lingvistic cuprins în atlasele dialectale:

> „Una dintre soluții ar putea fi, de exemplu, ca, din atlasele regionale, să fie excerptate și preluate toate datele care prezintă interes din punct de vedere lexicografic (cuvinte neatestate, sintagme, sensuri noi, variante fonetice și trăsături morfologice), în structuri logice ale articolelor consacrate cuvintelor-titlu, dar cu un sistem de trimitere simplificat pentru atestări (la indicele atlasului respectiv), urmând ca cel interesat să caute datele în discuție în volumele indicate, ce ar deveni *părți anexe* ale DLR" (Dumistrăcel et al. 2011: 234).

Iar întreg acest efort argumentativ al dialectologilor ieșeni vine ca urmare a convingerii conform căreia:

> „nu ar fi de acceptat ca dicționarul-tezaur al Academiei Române să nu răspundă, ca și până acum, celor mai pretențioase exigențe ale cunoașterii complexe a limbii române" (*ibidem*).

1.3. Importanța resurselor dialectale – cu preponderență a atlaselor lingvistice, dar și a glosarelor sau a dicționarelor de specialitate – pentru redactarea unor lucrări lexicografice complexe a fost reconfirmată recent în cadrul a două proiecte de cercetare dedicate unor terminologii de specialitate ale limbii române[4]. Dat fiind

[3] „Este evident sporul de informaţii pe care îl aduc atlasele regionale faţă de atlasul naţional, atât în ceea ce priveşte faptele dialectale, cât şi ariile diferitelor fenomene lingvistice" (Dumistrăcel et al. 2011: 229).

[4] Cele două proiecte, coordonate de CS I dr. Cristina Florescu de la Institutul de Filologie Română „Alexandru Philippide" al Academiei Române – Filiala Iaşi, au titlurile: *Terminologia românească meteorologică* (*ştiinţific* vs *popular*) *a stărilor atmosferice. Studiu lingvistic* (TMFA: 2011–2016) şi *Terminologia astronomică românească: ştiinţific* vs *popular. Fenomene, obiecte cosmice şi constelaţii* (TAFOC: 2021–2023).

interesul central acordat atât componentei științifice, cât și celei populare ale terminologiilor analizate – aspect care se poate observa chiar din titluri –, constituirea bazei de date necesară redactării lucrărilor științifice ce aveau să rezulte din aceste proiecte[5] a presupus o amplă documentare pe baza materialului dialectal din cele două serii ale atlaselor lingvistice românești: ALR și ALRR/NALR, la care s-au adăugat și dicționarele și glosarele dialectale publicate până acum. Comparativ cu situația existentă în cele două serii ale dicționarului-tezaur (DA/DLR) – ca punct de plecare în activitatea de documentare a cercetătorilor lexicografi din cele două echipe –, analiza materialului astfel obținut a avut ca principale rezultate:

– reconfirmarea unor atestări sau a unor arii dialectale mai vechi;

– creșterea numărului de atestări pentru anumiți termeni sau anumite forme și sensuri;

– atestarea unor noi termeni, sintagme, variante sau sensuri;

– facilitarea găsirii unor etimologii.

Situația care explică în bună măsură obținerea acestor rezultate are la bază decalajul existent între momentul publicării volumelor din seriile DA/DLR ale dicționarului-tezaur și apariția volumelor din seriile ALR și ALRR/NALR ale atlaselor lingvistice românești. Astfel, în timp ce fasciculele din seria DA au apărut în perioada 1913–1949, cele trei volume cu hărți analitice din ALR au văzut lumina tiparului în intervalul 1938–1942, motiv pentru care nu au putut fi folosite decât într-un mod limitat în activitatea de redactare a volumelor din seria DA. În ceea ce privește materialul dialectal arhivat și publicat de seria atlaselor lingvistice regionale (ALRR/NALR), acesta nu a putut intra în circuitul DLR decât odată cu apariția primelor volume din seria menționată (NALR–Olt., vol. I, 1967, ALRR–Mar., vol. I, 1969), existând, în mod firesc, o condiționare a dimensiunii arhivei de atestări dialectale pentru fiecare volum în parte din seria DLR prin situația publicării efective a atlaselor regionale la momentul efectuării operației de documentare pentru aceste volume.

Vom ilustra în continuare rezultatele menționate mai sus prin câteva exemple extrase din analizele de tip diatopic subsumate celor două proiecte de cercetare derulate în centrul academic ieșean[6].

1.3.1. Una dintre cele mai importante contribuții pe care lucrările de geografie lingvistică o pot aduce cercetărilor lexicografice ține de atestarea pentru prima dată a unor termeni/sintagme sau sensuri. Din cele câteva zeci de astfel de prime atestări înregistrate în cele două proiecte în urma consultării atlaselor, dicționarelor și glosarelor dialectale, redăm mai jos doar câteva:

– termeni/sintagme: *bujálă*[7] ‘arșiță’ (ALR II s.n. 1502, pct. 219, 260, 362; NALR–Mold., Bucov. 503, pct. 461, 462, 463, 478, 490, 493, 495; ALRR–Trans. 1563, pct. 245–247, 249–257, 262–269, 275, 277); *cocăiálă* ‘arșiță’ (NALR–Mold., Bucov. 1563, pct. 653, 655); *crăciunár* ‘decembrie’ (NALR–Olt. MN 273 [1556], pct. 986);

[5] Redăm aici volumele publicate ca urmare a cercetărilor realizate în intervalul 2011–2023: Cristina Florescu (coord.) 2015; Cristina Florescu (coord.) 2017; Cristina Florescu (coord.) 2023/2024.

[6] Analizele detaliate ale aspectelor menționate aici se regăsesc în Olariu 2015 și Olariu 2023/2024.

[7] Aflată în nord–nord-estul arealului dacoromânesc, aria de răspândire a acestui termen atestat acum pentru prima dată ne-a ajutat inclusiv în stabilirea etimologiei: v.sl. *bújati* (*бу́жать*) ‘a-și da sufletul, a-și da duhul’ (Vasmer 1964–1973 s.v.) (vezi discuția în Olariu 2015: 203).

a se crepeți (de ziuă) ‘a se ivi zorile’ (ALR II s.n. 762, pct. 64; NALR–Criș. 761, pct. 123; ALRR–Trans. MN 164–165 [1543], pct. 348, 349, 350, 373), *crăpui* ‘zori (de zi)’ (ALRR–Munt., Dobr. 504, pct. 831); *a se răzvăna (de ziuă)* ‘a se crăpa de ziuă’ (ALR II s.n. 760, pct. 574; ALRR–Trans. MN 164–165 [1543], pct. 317); *a se sâmcera/sâmceola (de ziuă)* ‘a se ivi zorile’ (NALR-Olt. 308, pct. 915, 932); *țurlúc* ‘țurțur’ (NALR–Mold., Bucov. 1615, pct. 530, 553, 600); *zorilă* ‘luceafărul de dimineață’ (ALRR-Munt., Dobr. 521, pct. 824, 827) etc.;

– sensuri: *fugău* ‘stea căzătoare’ (ALR II s.n. 807, pct. 362); *ploier* ‘noiembrie’ (NALR–Olt. MN 272 [1548], pct. 985); *pomar* ‘septembrie’ (ALRR–Munt., Dobr. 511, pct. 767; NALR–Mold., Bucov. 500, pct. 523); *sorb* „curcubeu” (ALM 44, pct. 215, 219); *a se zgâi (de ziuă)* ‘a se crăpa de ziuă’ (NALR–Criș. 761, pct. 186; ALRR–Trans. MN pl. 143–144 [1543], pct. 369); *a se ziui* ‘a se ivi zorile’ (NALR–Ban. 512, pct. 90) etc.

1.3.2. Apariția volumelor din seria ALRR/NALR a contribuit, pe lângă atestarea unor noi termeni și sensuri, și la reconfirmarea unor atestări sau a unor arii dialectale mai vechi, deja existente/configurate în lucrările lexicografice sau de cartografie lingvistică publicate. Un exemplu în acest sens îl reprezintă situația sintagmei *steaua ciobanului*, cu sensul de ‘luceafărul de seară’, care în DLR are ca atestări la nivel dialectal o singură ocurență din ALR: pct. 537 din centrul Moldovei. În urma documentării materialului din NALR–Mold., Bucov., am putut constata că respectivul punct era doar „epicentrul” unei zone mult mai întinse de actualizare a sintagmei în discuție, în total 20 de ocurențe (NALR–Mold., Bucov. 519, pct. 480, 486, 492, 493, 494, 535, 537, 548, 554, 556, 560, 562, 564, 565, 566, 573, 609, 619, 622, 624):

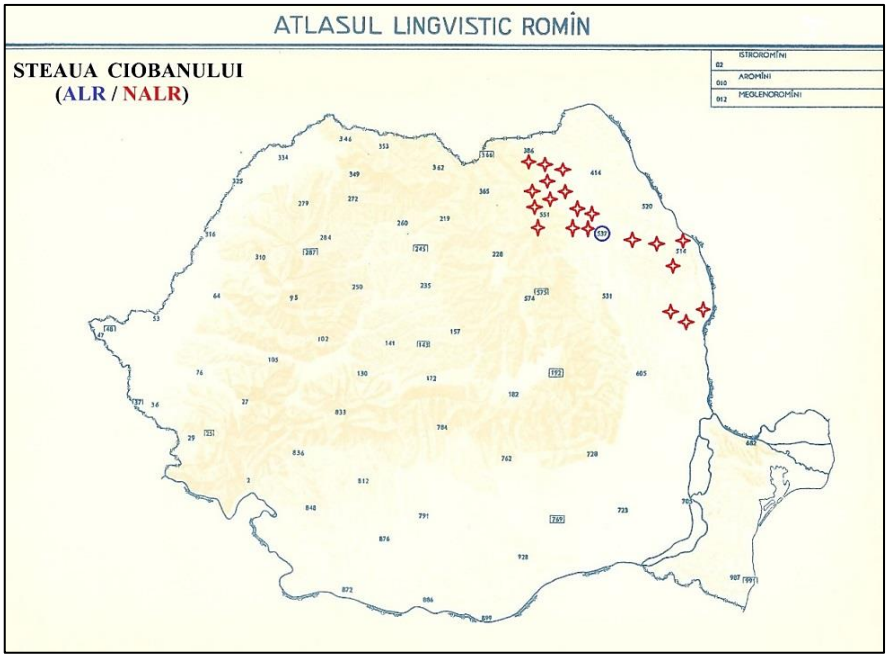

Figura 1. Prezența sintagmei *steaua ciobanului* ‘luceafărul de seară’ în ALR și ALRR/NALR.

Un caz interesant este cel al sintagmei *a se guri/gura de ziuă*, cu sensul de ‘a se ivi zorile, a se crăpa de ziuă’: cu o singură atestare în DA, sintagma nu se regăsește pe hărțile din ALR, însă apare în seria ALRR/NALR cu 19 ocurențe într-un areal ce include nu mai puțin de patru regiuni istorice (NALR–Ban. 512, pct. 16, 22, 25, 89, 91, 94; NALR–Olt. 308, pct. 901, 904, 909, 915; ALRR–Munt.Dobr. 504, pct. 671, 672, 678; ALRR–Trans. MN 164–165 [1543], pct. 429, 431, 432, 447, 448, 449), situație care susține importanța atlaselor regionale în identificarea unor fapte de limbă cu arii mai restrânse, ale căror ocurențe riscă adesea să treacă prin ochiurile rețelei atlasului național, fatalmente mai rară (vezi în acest sens Dumistrăcel et al. 2011). O altă semnificație care poate fi atribuită diferențelor între cele două serii de atlase în ceea ce privește această realitate lingvistică este supoziția unei extinderi relativ recente a ariei construcției *a se guri/gura de ziuă*. În sprijinul acestei ipoteze vin și situațiile întâlnite în DA, unde pentru *a se guri* cu sensul de ‘a se crăpa de ziuă’, ‘a se revărsa zorile’ avem doar o singură atestare, sau în dicționarul lui A. Scriban (a cărui perioadă de redactare se suprapune în linii mari cu perioada elaborării *Atlasului lingvistic român*), unde construcția nu apare.

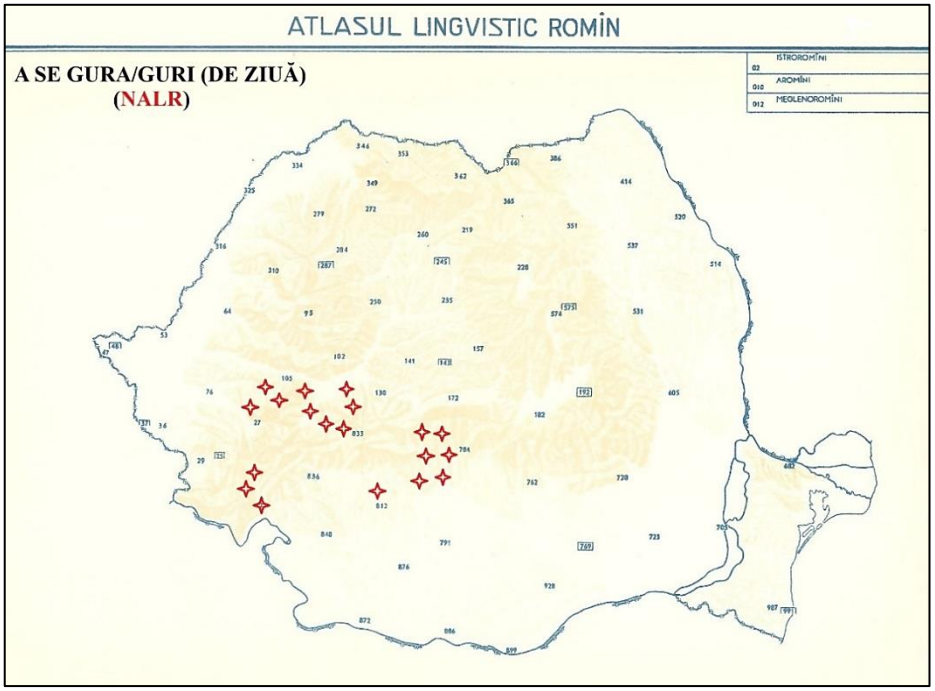

Figura 2. Prezența sintagmei *a se guri/gura (de ziuă)* în ALRR/NALR.

În sfârșit, dacă în cele două exemple anterioare am ilustrat maniera în care cuvinte/ sintagme prezente în seriile DA/DLR ale dicționarului-tezaur al limbii române sunt reconfirmate prin atestări în ALR și/sau ALRR/NALR, ultimul exemplu prezentat în discuția de față aduce în prim-plan rolul atlaselor lingvistice de a atesta (ALR) și apoi de a reconfirma (ALRR/NALR) anumiți termeni care nu se regăsesc

în dicționarul-tezaur al limbii române. Un asemenea caz este cel al sintagmei *a se gena/geni de ziuă*, cu același sens de ‘a se ivi zorile, a se crăpa de ziuă’, care, deși nu este atestată în DA, apare în două puncte din centrul Moldovei pe hărțile din ALR (ALR SN 760/762, pct. 514, 537), zonă reconfirmată, dar și mult extinsă apoi în rețeaua de puncte din NALR (NALR–Mold., Bucov. 494, pct. 482, 531, 536, 537, 540, 548, 550, 560, 567, 568, 582, 583, 584, 587, 597, 598, 599, 600, 601, 604, 605, 606, 609, 610, 612, 614, 617, 618, 623, 624, 626, 648):

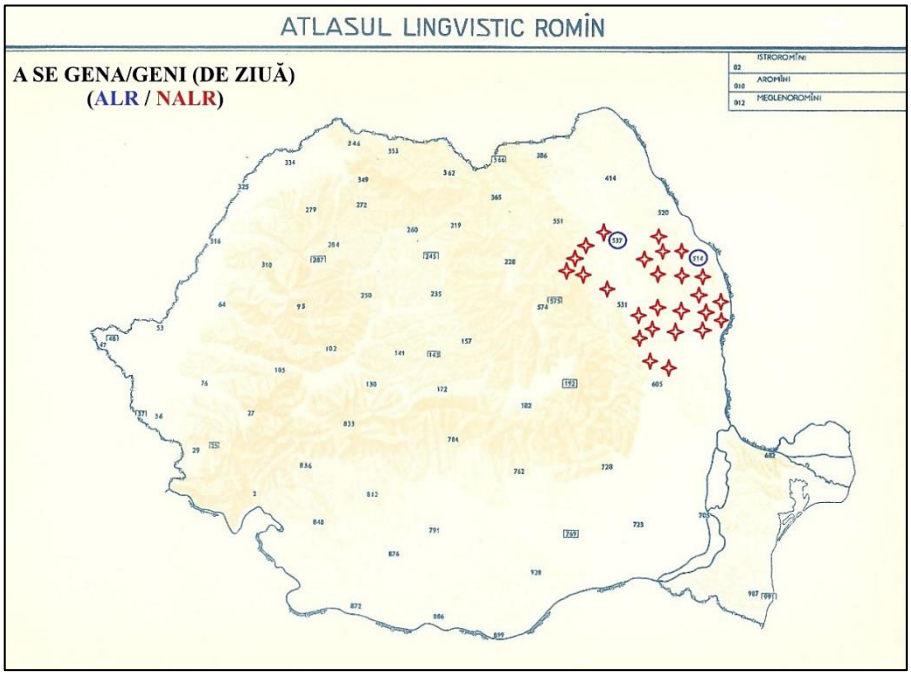

Figura 3. Prezența sintagmei *a se gena/geni (de ziuă)* în ALR și ALRR/NALR.

Din cele prezentate până acum putem observa importanța majoră a materialului lingvistic cuprins în atlasele naționale și în cele regionale pentru lucrările lexicografice dedicate limbii române. Din acest motiv, documentarea informației existente în aceste lucrări de geografie lingvistică a reprezentat în permanență o etapă esențială în activitatea redactorilor dicționarului-tezaur, care se desfășura printr-o consultare „clasică” a atlaselor și a volumelor de texte dialectale publicate, la rândul lor, în format „clasic”, pe hârtie. Însă, odată cu apariția în ultimele aproximativ două decenii a atlaselor editate sau chiar configurate *ab initio* în format digital, această activitate de documentare se reconfigurează, prin posibilitatea preluării şi utilizării de către lexicografi a instrumenetelor dezvolate în cadrul cercetărilor actuale din domeniul *digital humanities*. În partea a doua a lucrării vom prezenta principalele rezultate obținute în centrul academic ieșean în domeniul geografiei lingvistice digitale, reliefând inclusiv modalitatea în care aceste rezultate pot contribui la obținerea mult mai facilă a informației dialectale cuprinse în atlase și în volumele de texte dialectale de către redactorii lexicografi implicați în elaborarea noii serii din DLR.

2. ATLASE LINGVISTICE ÎN FORMAT DIGITAL

Investigațiile dedicate dezvoltării de sisteme informatice pentru digitalizarea și publicarea atlaselor lingvistice au început în urmă cu 25 de ani în centrele academice din Iași și Cluj-Napoca. Inițial, cercetările au urmărit realizarea de instrumente care să permită editarea asistată de calculator a planșelor din atlasele lingvistice regionale, iar ulterior au fost extinse pentru publicarea online a conținutului acestora și editarea textelor dialectale. Astfel, cercetătorii din cadrul Institutului de Informatică Teoretică (IITI) au elaborat, în colaborare cu dialectologii de la Institutul de Filologie Română „Alexandru Philippide" (IFRI) – ambele din cadrul Filialei Iaşi a Academiei Române –, aplicația *ALR-MB* utilizată pentru publicarea volumelor III, IV și V ale *Noului Atlas lingvistic român, pe regiuni*. *Moldova și Bucovina* (NALR–Mold., Bucov.) în anii 2007, 2014 și 2022. Pentru că, inițial, această aplicație a fost implementată și putea fi utilizată doar pentru planșele NALR–Mold., Bucov., ea a fost reproiectată în cadrul unei noi colaborări cu Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti" al Academiei Române din Bucureşti (ILIR). Noua versiune, denumită *AlrMaps*, poate fi utilizată pentru elaborarea planșelor oricărui atlas lingvistic din spațiul românesc. Astfel, cu ajutorul ALRMaps au fost publicate vol. al II-lea din *Atlasul lingvistic al dialectului aromân* (ALAR), *Atlasul lingvistic al dialectelor româneşti din nordul şi din sudul Dunării* (ALDRo) și este în curs de redactare vol. al VI-lea din NALR–Mold., Bucov. De asemenea, în cadrul unui proiect de colaborare România – Republica Moldova care este în curs de desfășurare, aplicația este extinsă împreună cu cercetătorii de la Universitatea de Stat a Moldovei (USM) pentru digitalizarea *Atlasului lingvistic moldovenesc* (ALM) (publicat inițial cu caractere chirilice) și republicarea acestuia folosind grafia latină. Pentru atlasele lingvistice care conțin doar hărți sintetice a fost elaborată aplicația *MapsGen*, utilizată pentru editarea *Atlas linguarum Europae* (ALE) vol. I, fasc. IX și a *Atlasului lingvistic român pe regiuni. Sinteză* (ALRR. *Sinteză*), vol. al IV-lea. În ceea ce privește publicarea online a resurselor dialectale, sunt disponibile versiunile interactive ale *Atlasului lingvistic audiovizual al Bucovinei* (ALAB) și ALDRo, în pregătire fiind eALM și eNALR–Mold., Bucov. (vezi *infra*, **2.4**). Nu în ultimul rând, trebuie amintite aplicațiile *AlrLibrary* (pentru editarea textelor dialectale, care dispune și de facilități multimedia pentru redarea sincronizată text – înregistrare audio) și *ALRWord AddIn*, componentă pentru editarea în Microsoft Word a textelor dialectale folosind transcrierea fonetică specifică limbii române.

2.1. Digitalizarea și publicarea planșelor atlaselor lingvistice regionale

Sistemul informatic *AlrMaps* este utilizat pentru digitalizarea informațiilor, generarea și editarea planșelor din atlasele lingvistice regionale. Componentele de bază ale sistemului sunt prezentate în partea stângă a figurii nr. 4. Baza de date geografice (BDG) conține descrierile pentru șabloanele hărților lingvistice, digitalizate cu ajutorul aplicației gratuite open-source QGIS de tip Sistem de Informații Geografice (Bejinariu et al. 2016). Șablonul include toate trăsăturile geografice, inclusiv punctele de anchetă ce vor fi relaționate cu transcrierile răspunsurilor din baza de date lingvistice (BDL).

Folosind interfața prietenoasă a aplicației *AlrMaps* (Fig. 5), utilizatorul editează BDL care conține lista punctelor de anchetă și cuvintele-titlu.

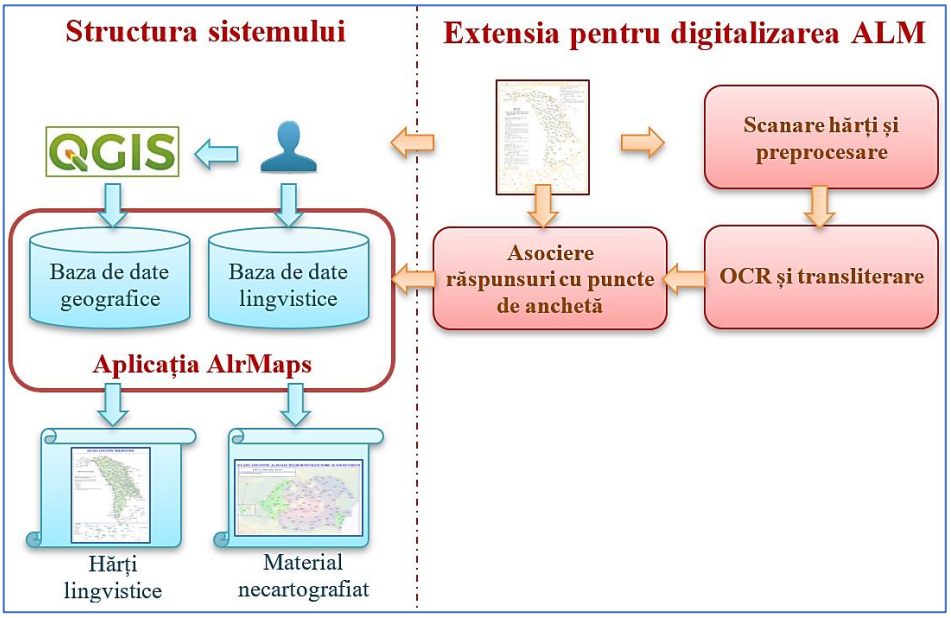

Figura 4. Structura sistemului *AlrMaps*.

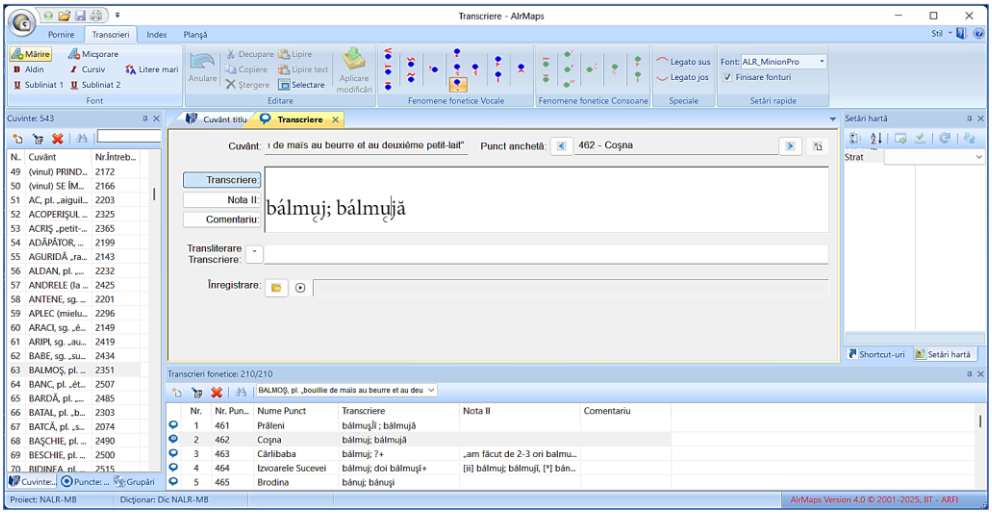

Figura 5. Interfața pentru editarea răspunsurilor în transcriere fonetică.

Lista punctelor de anchetă conține informații complete pentru fiecare dintre acestea: numărul punctului, numerele asociate din vechile atlase lingvistice (WLAD și ALR), numele localității, al comunei sau al orașului, județul, regiunea, țara, subdialectul folosit etc. Lista cuvintelor-titlu conține: cuvântul propriu-zis, numărul

întrebării și întrebarea pentru care cuvântul reprezintă un răspuns, eventuale traduceri în limbi străine, note explicative, o imagine asociată care va fi eventual afișată pe hartă, precum și transcrierile fonetice ale răspunsului, comentariului și notei explicative din fiecare punct de anchetă. Răspunsurile obținute în punctele de anchetă pot fi grupate manual sau automat, în funcție de similaritatea acestora sau de prezența anumitor forme fonetice. În plus, pentru fiecare dintre grupe se pot asocia culori, moduri de umplere și/sau simboluri în vederea redactării hărților sintetice sau a planșelor cu material necartografiat. Folosind informațiile stocate în cele două baze de date, aplicația *AlrMaps* generează automat planșele, care pot fi apoi editate (repoziționare, modificare atribute de desenare sau introducere de noi elemente grafice), tipărite și/sau salvate in format imagine (Fig. 6 și 7).

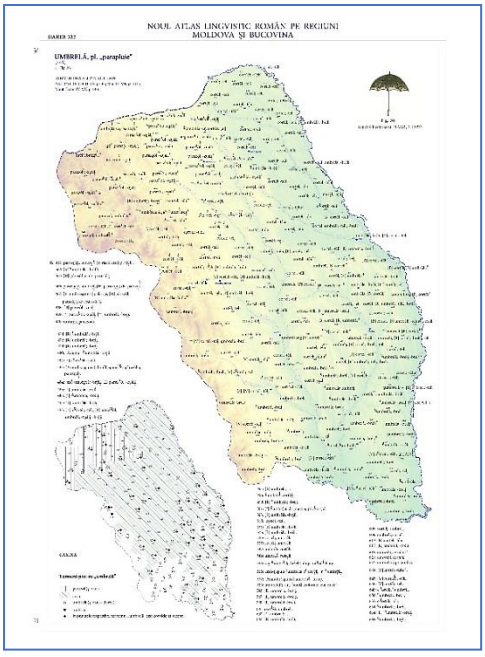

Figura 6. Planșă mixtă cu hartă analitică și hartă sintetică.

Figura 7. Planșă cu material necartografiat.

2.2. Extensia *AlrMaps* pentru publicarea ALM cu grafie latină

Unul dintre obiectivele principale ale proiectului PHILORD, care se derulează în perioada 2024–2026 sub forma unei colaborări între două echipe de lingviști și informaticieni din Iași și Chișinău, este republicarea cu caractere latine a *Atlasului lingvistic moldovenesc* (ALM), apărut în patru volume, în perioada 1968–1973, cu caractere chirilice. Prima soluție la care s-a recurs în cadrul proiectului – relativ cronofagă – constă în introducerea manuală a răspunsurilor transliterate în BDL. A doua soluție se bazează pe transliterarea automată și necesită implementarea unei extensii a aplicației *AlrMaps* (partea dreaptă a figurii nr. 4). Având în vedere vechimea

planșelor din ALM, într-o primă etapă acestea sunt scanate la rezoluție mare și preprocesate pentru a fi îmbunătățite calitativ (corecții de culoare, eliminare zgomot etc.). Apoi este aplicată o procedură de recunoaștere optică a caracterelor (OCR), folosindu-se modele antrenate special pentru caracterele chirilice. După o verificare preliminară a rezultatului recunoașterii automate este aplicată transliterarea automată în alfabetul latin. Este utilizat un set de 273 de reguli deja definite (Popovschi et al. 2022; Bumbu et al. 2023), câteva exemple fiind prezentate în tabelul I:

Tabelul I. Exemple de reguli utilizate în procesul de transliterare (Popovschi et al. 2022)

Transliteration rules			**Text sample**
Cyrillic alphabet		**Latin alphabet**	
у́	→	ú	пу́ни → púni (put on...)
ў	→	u̯	брыў → brîu (belt)
ʸоа́	→	ᵘo̯á	сʸоа́рили → sᵘo̯árili (sun)
уы	→	u̯â	доуы → do̯uâ (two)
ʸ + vowel	→	ᵘ + vowel	ʸогра́ды → ᵘográdâ (courtyard)

Rezultatul conține mai multe tipuri de informații care sunt prelucrate în mod diferit: răspunsurile de pe hartă, numerele punctelor de anchetă și notele aferente fiecărei hărți. Acestea sunt apoi importate automat în BDL după asocierea grafică a răspunsurilor cu punctele de anchetă. În continuare, după o ultimă verificare a rezultatelor transliterării, *AlrMaps* este utilizată pentru generarea, editarea și publicarea planșelor din ALM.

2.3. Indexarea informațiilor din atlasele lingvistice

Pe lângă planșele lingvistice sau cu material necartografiat, în atlase sunt incluse și diferite liste sau indexuri ale cuvintelor și punctelor de anchetă. Aplicația *AlrMaps* permite generarea automată a acestor liste în format tabelar în documente Microsoft Word sau fișiere HTML cu un format prestabilit, dar care poate fi modificat la nevoie. De asemenea, pot fi generate automat și liste ale răspunsurilor obținute pentru fiecare cuvânt în parte, care pot fi utile în etapa de verificare a transcrierilor fonetice. Un instrument util este căutarea aparițiilor unui anumit cuvânt sau forme/secvențe fonetice în întregul dicționar. Cu ajutorul interfeței din *AlrMaps* pot fi inițiate operații de căutare simplă, în care sunt identificate potrivirile exacte ale unui cuvânt sau forme specificate de utilizator. Sunt disponibile și instrumente de căutare complexă cu specificarea poziției formei (prefix sau sufix), a prezenței unor fenomene fonetice sau a unor poziționări ale simbolurilor, la nivel de cuvânt sau la nivelul întregului dicționar. Și în acest caz, rezultatele obținute pot fi salvate ca document Word într-un format tabelar prestabilit (Fig. 8). În cazul căutării după o formă prestabilită la nivelul dicționarului pot fi generate și hărți sintetice în care figurează frecvența de apariție în fiecare punct de anchetă (Fig. 9). Numărul de intervale și culorile utilizate pe hartă pentru fiecare interval sunt stabilite automat de aplicație, dar pot fi particularizate de utilizator după preferință (Bejinariu et al. 2024).

Proiect: Dic NALR-MB

Index formă: ţa

Filtru cuvânt: -
Filtru punct: -
Căutare în: Răspuns;
Căutare ca: Transcriere;
Alte opţiuni: *** Opţiuni detaliate;

Nr.crt.	Cuvânt		Punct de anchetă		Răspuns	Nota II	Comentariu
1	[2301]	(cum) ALUNGI (oaia)? „comment repoussez-vous un mouton?”	501	Dumbrăveni	nà-nà-nà, u̯oi̯ița, nà-nà-nà!		
2		-//-	651	Câmpineanca (Floreşti)	ţa, u̯ăi̯i!	[*] bĭr!	
3	[2300]	(cum) CHEMI (oaia)? „commnet faîtes-vous s’approcher un mouton?”	468	Deluţ	ţ̆rița, Bîrúța, ţ̆rița!		
4		-//-	511	Rădăuţi-Prut	brița-brița!, na-na-na!		
5		-//-	535	Grozeşti	na, u̯oița, na-na-na!		
6		-//-	569	Cracăul Negru	na-na-na, oi̯ița!		
7		-//-	592	Măgura	sugurę!, na, u̯oița, nà!		
8		-//-	599	Negreşti	na, Stelúța!		
9		-//-	609	Micleşti	na-na, u̯oi̯ița!		
10		-//-	618	Lipovăţ	na, u̯oița, na!		
11		-//-	622	Stănileşti	na-na-na, u̯oița, na!		
12		-//-	634	Vârlezi	na, u̯oița, na-na-na!		

1

Figura 8. Lista aparițiilor unei anumite forme/secvenţe fonetice generată ca document Word.

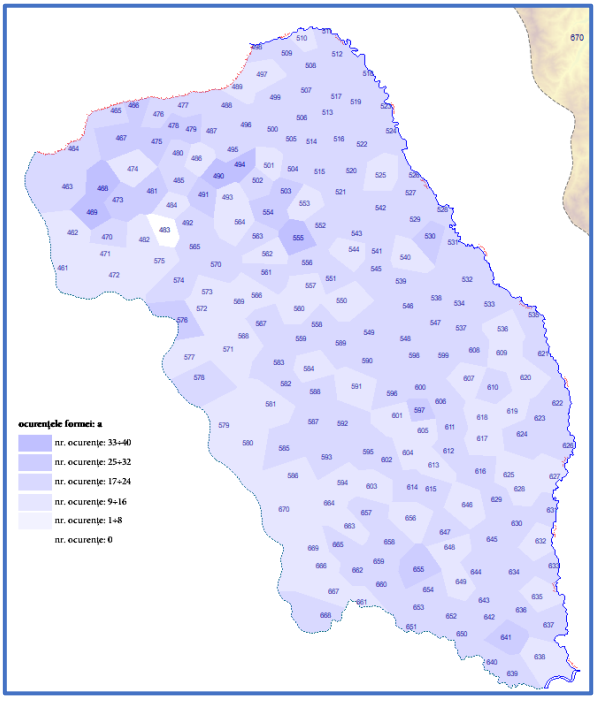

Figura 9. Reprezentarea grafică a numărului de ocurenţe pentru un fenomen lingvistic pe harta sintetică.

2.4. Atlase lingvistice online

Toate instrumentele prezentate în secțiunile anterioare permit consultarea și publicarea atlaselor lingvistice în format tipărit. Au fost dezvoltate însă și versiuni online care permit accesul public la resursele dialectale. Astfel, *Atlasul lingvistic audiovizual al Bucovinei* (ALAB), disponibil la adresa https://www.philippide.ro/alab/index.html, permite urmărirea atât a variaţiei diatopice, cât şi a celei distratice (variația diasexuală + variaţia diagenerațională), pentru unele puncte de anchetă fiind luată în considerare inclusiv variația etnolingvistică. În ALAB sunt incluse și etnotexte, sub forma unor înregistrări video având ca temă diferite obiceiuri calendaristice sau rituri de trecere (Olariu et al. 2016).

Un proiect mai amplu este eNALR-MB care își propune, ca prim obiectiv, publicarea online a celor cinci volume de atlas şi a celor trei volume cu etno- şi sociotexte tipărite până acum în cadrul programului NALR-Mold., Bucov. (Olariu et al. 2022, Olariu et al. 2024). În plus, acestea vor fi însoțite de o hartă sonoră cu înregistrări audio referitoare la diferite obiceiuri, activități şi întâmplări din viaţa subiecţilor. Fiecărui fișier audio îi sunt asociate informații despre teme, puncte de anchetă, informatori, clase de texte dialectale și limba vorbită. Interfața de administrare este implementată, dar deocamdată nu este publică, fiind în curs de testare și încărcare cu informații (Fig. 10).

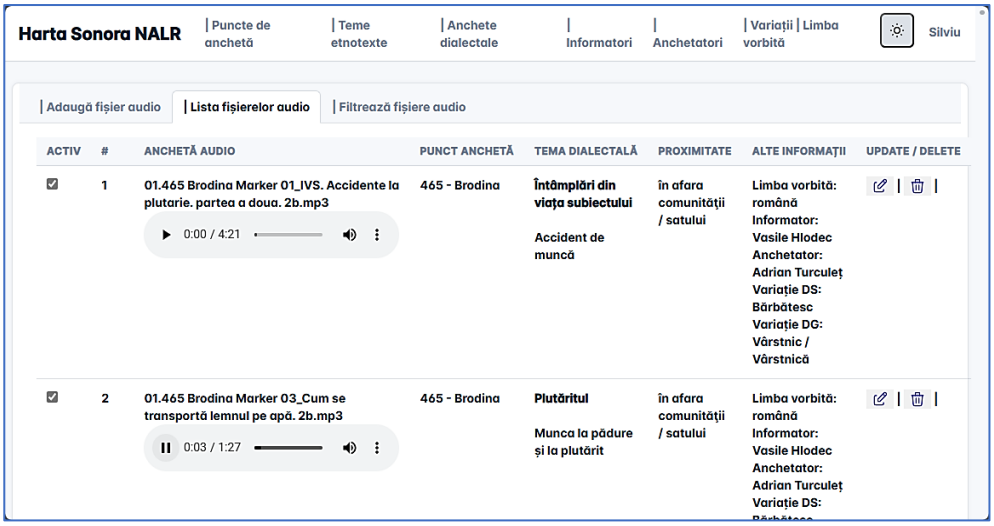

Figura 10. Interfața web pentru consultarea arhivei audio.

2.5. Editarea textelor dialectale

Textele orale în registru dialectal [= „texte dialectale”] reprezintă, alături de volumele de atlas, cea de-a doua componentă majoră a proiectelor ALRR/NALR, pentru a căror digitalizare s-au derulat cercetări susținute în ultima perioadă. Astfel, prin colaborarea cercetătorilor de la cele trei institute (IITI, IFRI şi ILIR) au fost

proiectate și implementate două aplicații pentru editarea textelor dialectale folosind transcrierea fonetică specifică limbii române (Bejinariu et al. 2022). Prima dintre acestea, deși pune la dispoziție funcții avansate de editare, folosește o codificare proprie a textului și, prin urmare, nu poate comunica cu alte editoare de text. Din acest motiv a fost dezvoltat un add-in pentru Microsoft Word, care să permită editarea de texte cu transcriere fonetică folosindu-se unul dintre cele mai cunoscute editoare de text. Interfața intuitivă (Fig. 11) permite selectarea variantelor de simboluri și aplicarea fenomenelor fonetice sau a accentelor pentru fiecare simbol de bază introdus din tastatură, cu ajutorul comenzilor disponibile într-un panou introdus pe bara de instrumente. Fiind integrat în Word, sunt disponibile toate funcțiile de editare avansată specifice acestui program.

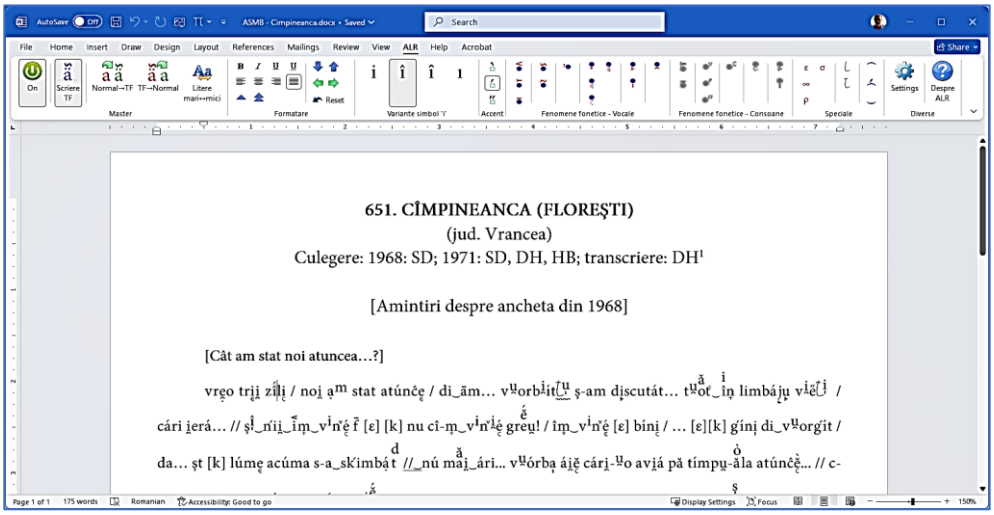
651. CÎMPINEANCA (FLOREŞTI)
(jud. Vrancea)
Culegere: 1968: SD; 1971: SD, DH, HB; transcriere: DH¹
[Amintiri despre ancheta din 1968]
[Cât am stat noi atuncea…?]

Figura 11. Editarea textelor dialectale în Microsoft Word.

3. CONCLUZII

Implicarea în ultimele două decenii a noilor tehnologii informaționale în cercetările de geolingvistică românească a condus la crearea de către echipe mixte de lingvişti şi informaticieni din principalele centre academice din România (Iaşi, Cluj-Napoca, Bucureşti) a unor aplicații informatice capabile să contribuie atât la editarea în format electronic a principalelor instrumente de cercetare specifice domeniului (în primul rând a atlaselor lingvistice, dar şi a volumelor de texte dialectale), valorificându-se astfel materialul lingvistic încă nepublicat din arhivele atlaselor regionale, cât şi la crearea *ab initio* în format digital a unor astfel de platforme cu material dialectal. Unele dintre principalele avantaje pentru cercetarea lingvistică (implicit lexicografică) ale acestor proiecte de digitalizare sunt date, în primul rând, de accesul facil şi imediat la informaţie, ca urmare a formatului online, dar şi de faptul că ele includ anumite funcţii specifice care permit interogarea interactivă şi focalizată a materialului lingvistic (ca urmare a existenţei unor baze de date) sau structurarea

informaţiei sub formă de indexuri, în conformitate cu anumiţi parametri de căutare specificaţi de utilizator. Toate aceste facilităţi generate de integrarea tehnologiilor digitale în realizarea unor proiecte – noi sau deja în derulare – de cartografie lingvistică duc, în cele din urmă, la o mult mai bună cooptare şi, în acelaşi timp, valorificare a resurselor dialectale în analizele dedicate limbii române, fie acestea sub forma unor studii mai aplicate sau chiar a unor lucrări fundamentale, aşa cum este şi dicţionarul-tezaur. Iar aceste noi realităţi epistemice nu fac altceva decât să confirme ceea ce acelaşi Sextil Puşcariu – invocat în partea de început a prezentei lucrări – afirma cu aproape un secol în urmă:

> „atlasul este o comoară de informații pentru cel ce știe să-l utilizeze și un izvor nesecat de probleme nouă. Precum din volumul de poezii al unui poet mare fiecare critic citează alte versuri spre a arăta frumusețile operei, tot astfel fiecare lingvist va scoate din hărțile Atlasului alt material prețios pentru preocupările sale științifice” (Pușcariu 1938: 405–406).

Parafrazând ideile lui Puşcariu, suntem convinşi că aplicaţiile informatice cu ajutorul cărora sunt concepute şi editate la ora actuală atlasele dialectale îi vor ajuta pe lingvişti, pe de o parte, să optimizeze utilizarea „comorii de informaţii” cuprinsă în aceste lucrări, iar pe de altă parte, să-i stimuleze în direcţia formulării unor „probleme nouă”, de exemplu ca urmare a unor analize de tip cantitativ mult mai elaborate, pe care astfel de metode de cercetare bazate pe tehnologiile informaţionale actuale pot să le susţină.

BIBLIOGRAFIE

Bejinariu et al. 2016 = Silviu-Ioan Bejinariu, Ramona Luca, Florin-Teodor Olariu, *A GIS Based Approach for Information Management in Geolinguistics*, în „Memoirs of the Scientific Sections of the Romanian Academy”, XXXIX, Computer Science, p. 37–45.

Bejinariu et al. 2022 = Silviu-Ioan Bejinariu, Ramona Luca, Vasile Apopei, Florin Iftene, *Editarea textelor dialectale folosind transcrierea fonetică specifică limbii române*, în „Anuar de lingvistică și istorie literară”, LXII, p. 37–53.

Bejinariu et al. 2024 = Silviu-Ioan Bejinariu, Ludmila Malahov, Tudor Bumbu, Florin-Teodor Olariu, Ramona Luca, *New Achievements in the Digitization of Linguistic Atlases of the Romanian Language*, „The 19th International Conference on Linguistic Resources and Tools for Natural Language Processing – ConsILR-2024”, 14–16 noiembrie 2024, Alba Iulia.

Bumbu et al. 2023 = Tudor Bumbu, Lyudmila Burtseva, Svetlana Cojocaru, Alexandru Colesnicov, Ludmila Malahov, *Distinctive features of recognition for documents printed in the Romanian transitional alphabets*, în „Computer Science Journal of Moldova”, vol. 31, no. 3(93), p. 340–350.

Dumistrăcel et al. 2011 = Stelian Dumistrăcel, Doina Hreapcă, Luminița Botoșineanu, *De la atlasul lingvistic național la atlasele regionale: semnificația diferențelor*, în *Studii de dialectologie, istoria limbii și onomastică. Omagiu domnului Teofil Teaha*, volum coordonat de Maria Marin și Daniela Răuțu, București, Editura Academiei Române, p. 219–242.

Dumistrăcel et al. 2015 = Stelian Dumistrăcel, Doina Hreapcă, Luminița Botoșineanu, *Lucrările românești de geografie lingvistică și* ***Dicționarul limbii române*** *al Academiei*: *oferta lexicografică a materialului necartografiat și problema „literarizării”*, în „Anuar de lingvistică și istorie literară”, LV, p. 19–72.

Florescu (coord.) 2015 = Cristina Florescu (coord.), Laura Manea, Elena Tamba, Alina Pricop, Cristina Cărăbuș, Liviu Apostol, Florin-Teodor Olariu, Mădălin-Ionel Patrașcu, *Terminologia meteorologică românească a fenomenelor atmosferice* (*științific* versus *popular*), Iași, Editura Universității „Alexandru Ioan Cuza”.

Florescu (coord.) 2017 = Cristina Florescu (coord.), Laura Manea, Elena Tamba, Alina Pricop, Cristina Cărăbuș, Liviu Apostol, Florin-Teodor Olariu, Mădălin-Ionel Patrașcu, *Dicționarul fenomenelor atmosferice*, Iași, Editura Universității „Alexandru Ioan Cuza”.

Florescu (coord.) 2023/2024 = Cristina Florescu (coord.), Laura Manea, Elena Isabelle Tamba, Alina-Mihaela Bursuc, Daniela Butnaru, Cristina-Mariana Cărăbuș, Maria-Marilena Ciobanu, Florin-Teodor Olariu, Cătălin Galeș, Gabriela-Ana Azanfirei, Claudius Teodorescu, *Terminologia astronomică românească științifică și populară: fenomene, obiecte cosmice și constelații*, Cluj-Napoca, Presa Universitară Clujeană, 2023 (ediție revizuită: 2024).

Hreapcă 2012 = Doina Hreapcă, *Valorificarea lexicografică a cercetărilor românești de geografie lingvistică*, în „Anuar de lingvistică și istorie literară”, LII, p. 216–222.

Olariu 2015 = Florin-Teodor Olariu, *Aspecte dialectale*, în Cristina Florescu (coord.), *Terminologia meteorologică românească a fenomenelor atmosferice* (*științific* versus *popular*), Iași, Editura Universității „Alexandru Ioan Cuza”, p. 183–212.

Olariu et al. 2016 = Florin-Teodor Olariu, Veronica Olariu, Ramona Luca, *Atlasul lingvistic audiovizual al Bucovinei (ALAB) – repere metodologice*, în *Lucrările celui de-al XVI-lea Simpozion internațional de dialectologie*, volum editat de Dumitru Loșonți, Veronica Vlasin, Nicolae Mocanu, Cluj-Napoca, Argonaut & Scriptor, p. 327–329.

Olariu et al. 2022 = Florin-Teodor Olariu, Veronica Olariu, Silviu-Ioan Bejinariu, Ramona Luca, *NALR. Moldova și Bucovina – resurse lingvistice în format electronic (eNALR-MB)*, Atelierul de lucru „Geografia lingvistică românească în era digitalizării”, 21–22 octombrie 2022, Iași, Academia Română (onsite & online).

Olariu et al. 2024 = Florin-Teodor Olariu, Alexandru Laurențiu Cohal, Luminița Botoșineanu, Veronica Olariu, Ramona Luca, Silviu-Ioan Bejinariu, *Making Linguistic Resources Accessible. The Audio Archive of the New Romanian Linguistic Atlas by Regions. Moldova and Bukovina*, în *Proceedings of the 18th International Conference on Linguistic Resources and Tools for Natural Language Processing – ConsILR-2023, Brașov, 11–14 decembrie 2023*, volum editat de Anca Vasilescu, Daniela Gîfu, Dan Tufiș, Dan Cristea, Iași, Editura Universității „Alexandru Ioan Cuza” din Iași, p. 199–214.

Olariu 2023/2024 = Florin-Teodor Olariu, *Terminologia astronomică românească – aspecte dialectale*, în Cristina Florescu (coord.), *Terminologia astronomică românească științifică și populară: fenomene, obiecte cosmice și constelații*, Cluj-Napoca, Presa Universitară Clujeană, 2023, p. 195–225 (reeditat 2024).

Popovschi et al. 2022 = Liliana Popovschi, Ludmila Malahov, Vlada Colesnicova, *Digital processing of dialectal texts published in Chisinau in the years 1969–1987*, în „Akademos” 4(63)/2021, Chișinău, p. 48–53.

Pușcariu 1931 = Sextil Pușcariu, *Pe marginea cărților III*, în „Dacoromania”, VI, 1929–1930, p. 484–536.

Pușcariu 1938 = Sextil Pușcariu, *Pe marginea cărților VI*, în „Dacoromania”, IX, 1936–1938, p. 403–449.

Vasmer 1964–1973 = Max Vasmer, *Etimologiceskij slovar' russkogo jazyka*, vol. I–IV, Moskva.

LINGUISTIC ATLASES AS LEXICOGRAPHIC RESOURCES: FROM PRINTED EDITIONS TO DIGITAL FORMAT

ABSTRACT

This paper brings back into focus the major importance of the material contained in national and regional linguistic atlases for lexicographic works dedicated to the Romanian language. Starting from this undeniable reality, the documentation of the information existing in these linguistic geography works has always represented an essential stage in the work of the editors of the thesaurus dictionary, which was carried out through a "classical" consultation of atlases and volumes of dialectal texts, themselves published in "classic" printed format. However, with the emergence in the last two decades of atlases edited or even conceived *ab initio* in digital format, this documentation activity is being reconfigured, through the possibility of lexicographers taking over and using the tools developed by current research in the field of *digital humanities*. Some of the main advantages for linguistic (and implicitly lexicographic) research offered by these digitization projects are, first of all, the easy and immediate access to information due to the online format, as well as the inclusion of specific functions that allow interactive and targeted querying of linguistic material (as a result of the existence of databases), or structuring of information in the form of indexes, according to user-specified search parameters. All these facilities made possible by the integration of digital technologies into the development of linguistic cartography projects – new or already in progress – ultimately lead to a much better integration and, at the same time, valorization of dialectal resources in analyses dedicated to the Romanian language.

Keywords: *linguistic atlases*, *printed* vs *digital format*, *linguistic resources & databases*, *Geographic Information System (GIS)*, *Optical Character Recognition (OCR)*, *digital humanities*.